LEAD SCORING CASE STUDY USING LOGISTIC REGRESSION

By:

- Dheeraj Singh

Contents

Problem Statement

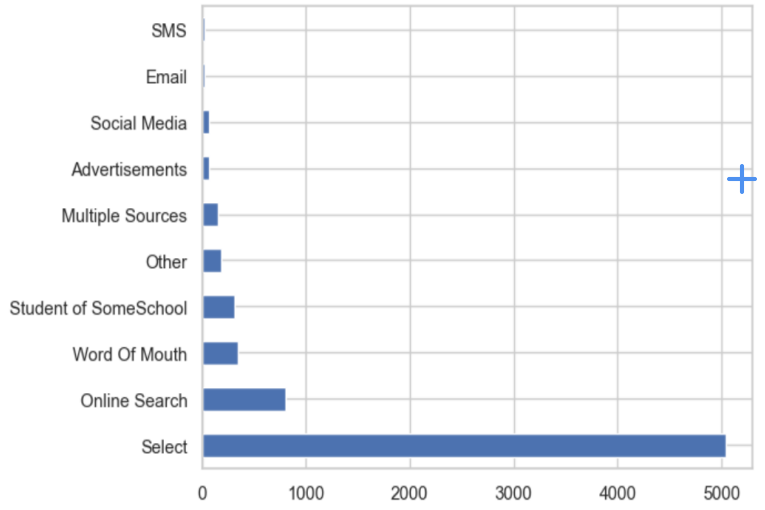
- An education company named X Education sells online courses to industry professionals. On any given day, many professionals who are interested in the courses land on their website and browse for courses. They have process of form filling on their website after which the company that individual as a lead.
- Once these leads are acquired, employees from the sales team start making calls, writing emails, etc.Through this process, some of the leads get converted while most do not.
- The typical lead conversion rate at X education is around 30%. Now, this means if, say, they acquire 100 leads in a day, only about 30 of them are converted. To make this process more efficient, the company wishes to identify the most potential leads, also known as Hot Leads.
- If they successfully identify this set of leads, the lead conversion rate should go up as the sales team will now be focusing more on communicating with the potential leads rather than making calls to everyone

Problem Approach

- Importing and inspecting the dataframe.
- Data Cleaning and preparation
- EDA
- Dummy variables creation
- Test-Train split
- Correlations
- Model Building
- Model Evaluation
- Making predictions on test set

EDA – Data Cleaning

- There are few irrelevant columns and some columns have data which is not useful during analysis, for example "Select" value is not useful but present in multiple columns.

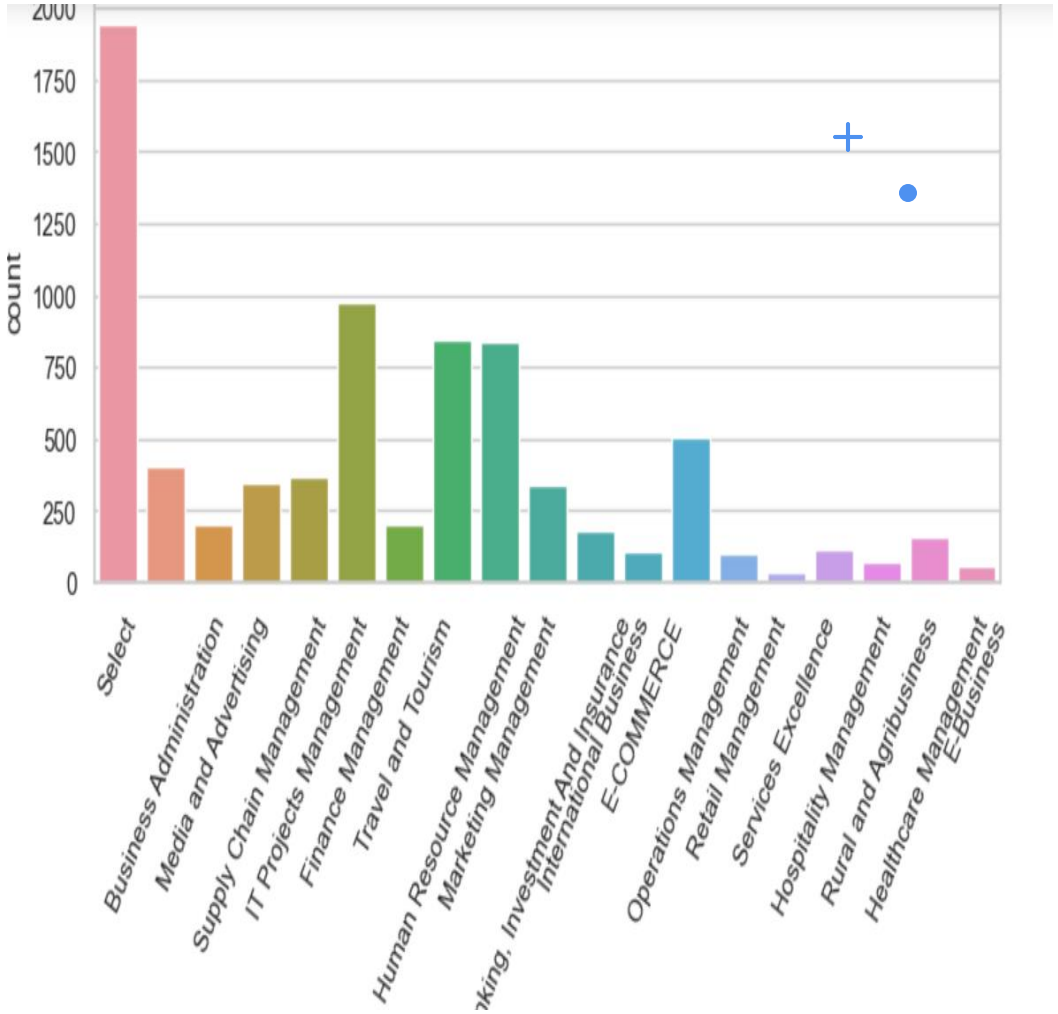



```
leads['Specialization'].value_counts().plot(kind='barh')
```

```
<AxesSubplot:>
```

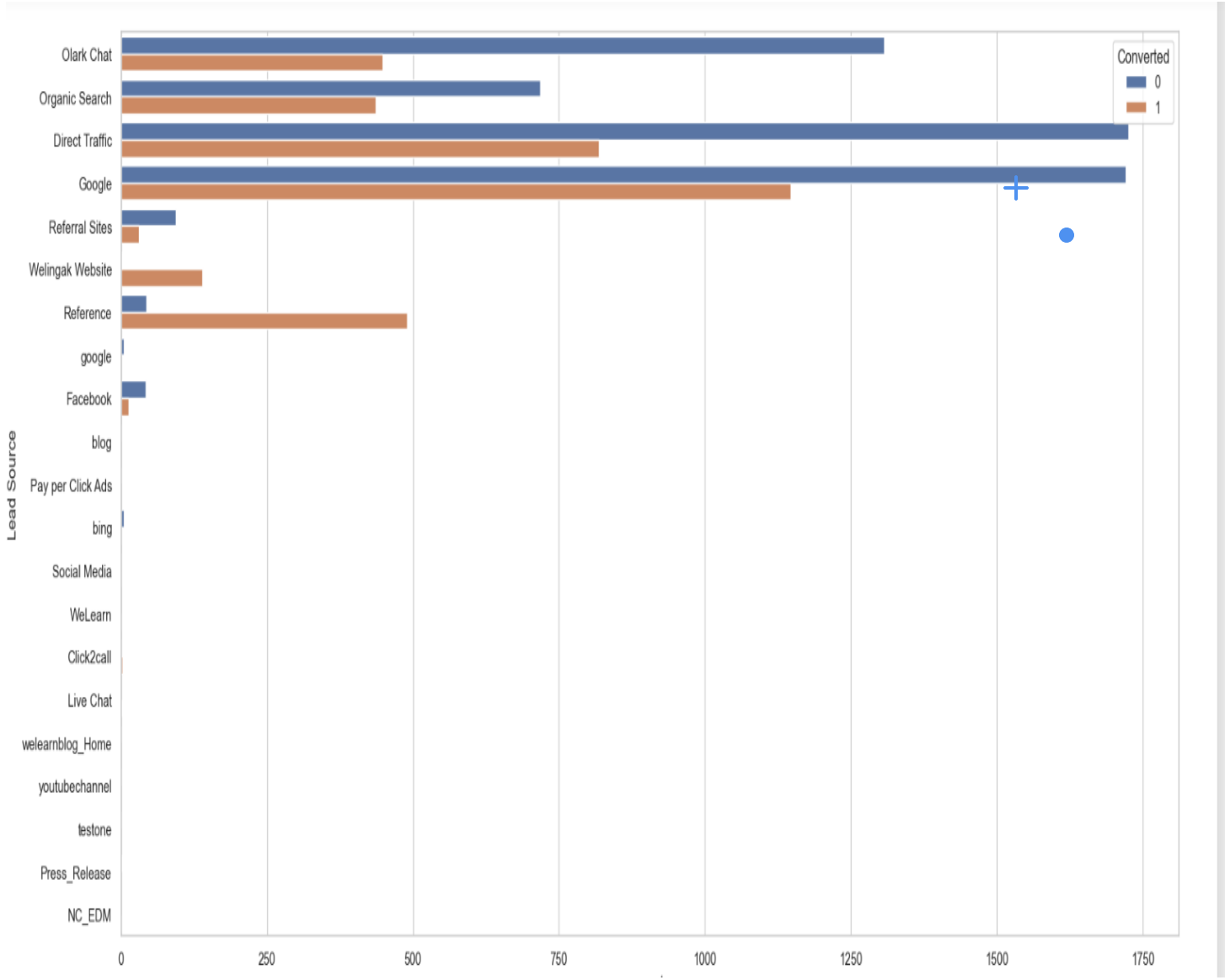
Specialization

- Leads from HR, Finance and Marketing Management are high probability to convert.



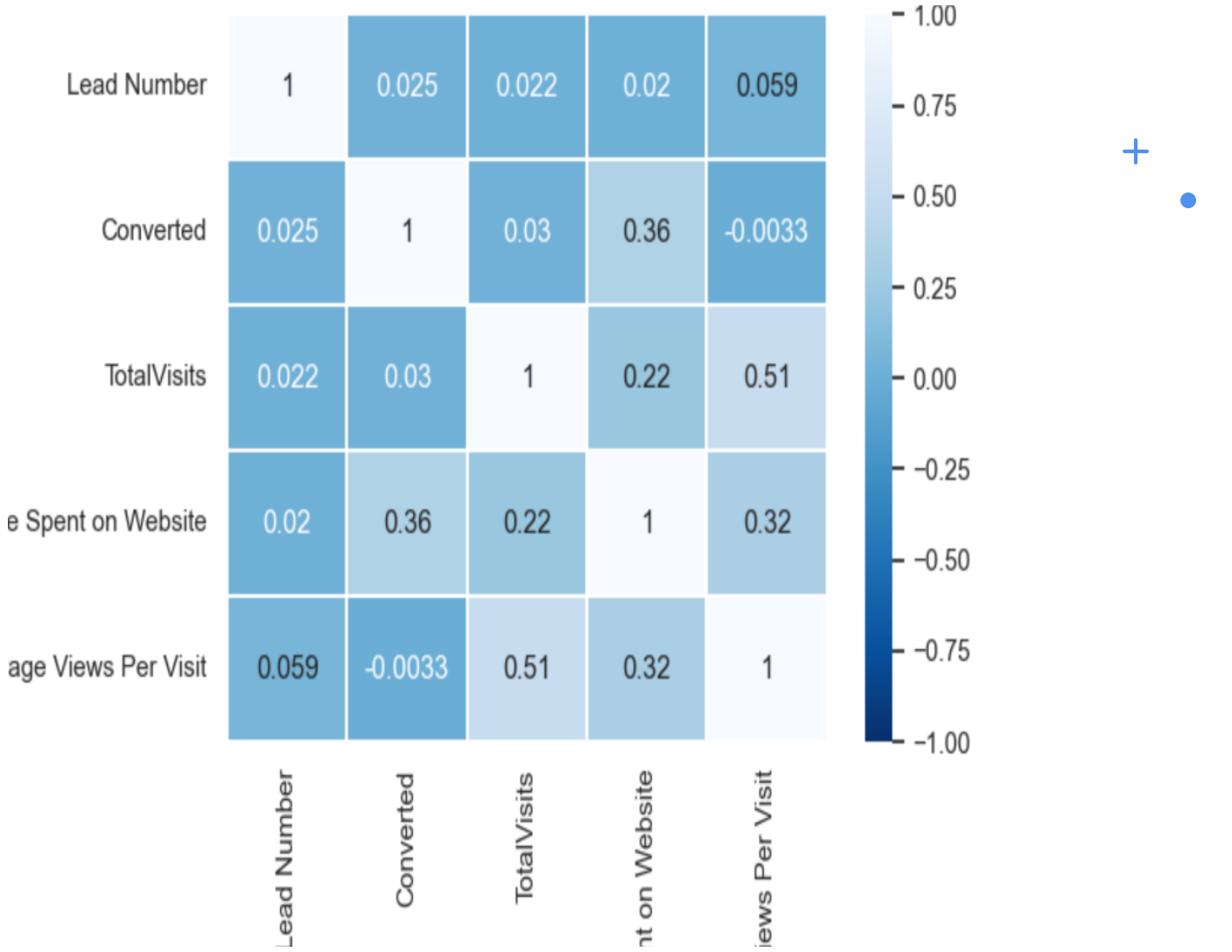
Lead Source

- Most of the leads are coming from Google, Direct Traffic and Olark chat. Abd most of them have got converted to business.

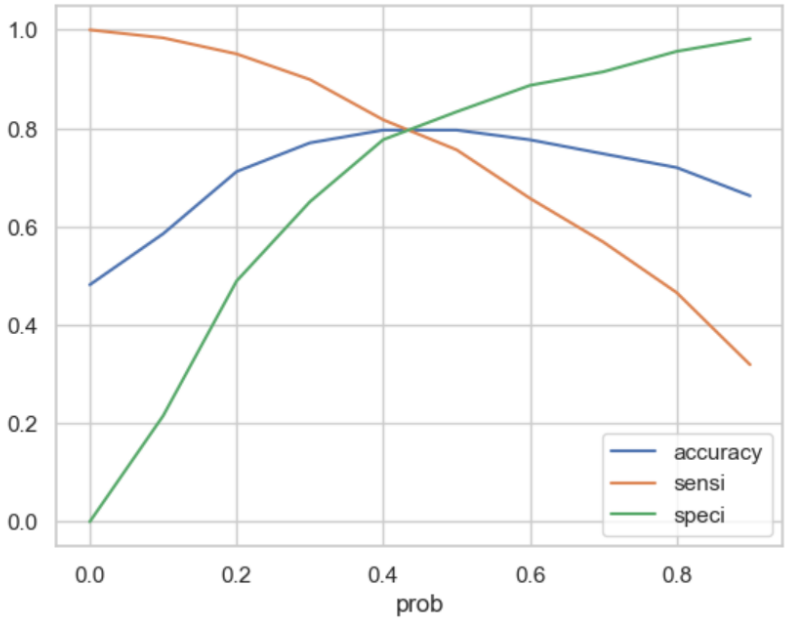


Correlation

- There is no correlation between the variables.



Model Evaluation (ROC curve)



- 0.43 is the tradeoff between Precision and Recall- So, we can consider Prospect Lead conversion rate is >43%.

Observations

- Train Data:

Accuracy: 80%

Sensitivity: 77%

Specificity: 80%

- Test Data:

Accuracy: 80%

Sensitivity: 77%

Specificity: 80%

- Feature List
  - Lead Source_Olark Chat
  - Specialization_Others
  - Lead Origin_Lead Add Form
  - Total Time Spent on Website
  - What is your current occupation_Working Professionals
  - Do Not Email

Conclusion

- We see max number of leads are generated by google / direct traffic. Max conversion ratio is by reference and welingak website.
- Leads who spent more time on website, more likely to convert.
- Most common last activity is email opened. highest rate = SMS Sent. Max are unemployed. Max conversion with working professional.